主成分分析による
景気指標のウエイトの算出と検定

勝　浦　正　樹

はじめに

景気動向指数（Diffusion Index；DI）に代表される景気指標[1]に関しては，これまでに様々な研究が行われ，問題点などが指摘されてきている。[2] その1つとして，ウェイトづけの問題がある。すなわち，各個別指標を総合化して（具体的にいうと平均して）景気指標を作成する際，各個別指標にウェイトをつける（加重平均をする）かどうか，またウェイトをつける場合，どのような方法でウェイトを算出したらよいか，といった問題である。

そこで本論文では，景気指標にウェイトはつけた方がよいという立場にたって，主成分分析（Principal Component Analysis）を用いたウェイトの算出に関する考察を行う。特に，主成分分析のもつ記述統計的な有効性からばかりでなく，主成分分析における統計的推論の，景気指標のウェイトの検定への適用可能性に目を向け検討してみる。

主成分分析は，Pearson [24] においてその基本概念が示された。その後，Hotelling によって「主成分」という名称が与えられ（Hotelling [12]），多変量解析の1手法として確立し，利用されてきている。とりわけ主成分分析は，心理学などの分野を中心に発達し，実証的に応用されている。[3] しかし，そうした実証分析においては，母集団と標本との関係を考慮しない記述統計の方法として主に利用されてきた。他方，主成分分析を，多変量解析における純粋に統計理論的な枠組の中でとらえ，推測統計の理論を適用しようとする動きも，数理統計学者を中心にみられるようになっている。[4]

以下では，まず主成分分析を概説し，景気指標のウェイトの算出方法としての有効性について論述する。さらに，主成分分析における統計的推論のなかで，固有値・固有ベクトルに関する検定を中心にまとめ，その意義・具体的な利用法・問題点などを考察する。最後に，これらの点をふまえた上で，実際に景気指標のウェイトの算出と検定に主成分分析を利用し，実証的な分析を行う。なお，ここで用いる景気指標は，最近注目を浴びているコンポジット・インデックス（Composite Index；以下 CI と略す）に限定する。

注(1) 景気指標とは，いくつかの個別の経済統計データ（個別指標）を何らかの方法で総合化し，景気を数量的に表現しようとするものである。なお，景気指数という言葉とは，若干のニュアンスの違いはあるにせよ，ほぼ同義に用いている。
(2) 刈屋 [15]，溝口・浜田 [21]，森 [22]，田原 [33]，Zarnowitz and Boschan [36] [37] などを参照のこと。
(3) 心理学以外にも，人類学，医学，地質学，物理学，生物学，社会学などにも利用されている。また，経済学への応用も試みられている。
(4) これらについては，Girshick [10] [11] 以来，Bartlett [3]，Lawley [18]，Kshirsagar [16]，Anderson [1]，Rao [25] などをはじめ，その他多くの論文が著されている。主成分分析の歴史的展開は，Kshirsagar [17] pp. 459–464 にくわしい。

1. 主成分分析と CI のウェイト

(1) 主成分分析概説

いま，p 個の変量 X_{1i}，X_{2i}……X_{pi} ($i=1$……N; N は観察期間）が与えられているとしよう。そこで，これらを線形的に結合させた

$$z_1=\beta_1X_{1i}+\beta_2X_{2i}+\cdots\cdots+\beta_pX_{pi}=\boldsymbol{\beta}'\boldsymbol{X} \tag{1.1}$$

という総合的指標を考える。ここで $\boldsymbol{\beta}'=[\beta_1\cdots\cdots\beta_p]$ は，$\boldsymbol{\beta}'\boldsymbol{\beta}=1$ という制約条件のもとで $z_1=\boldsymbol{\beta}'\boldsymbol{X}$ の分散が最大になるように決定される。この z_1 のことを Hotelling は，第1主成分と呼んだ。

具体的には，制約条件付の最大化問題として，Lagrange 未定乗数法により $\boldsymbol{\beta}$ は決定される。過程は省略し(5) 結果だけを示すと，主成分分析は結局，

$$\boldsymbol{\Sigma\beta}=\lambda_1\boldsymbol{\beta} \tag{1.2}$$

という固有値問題と等しくなる。ここで，$\boldsymbol{\Sigma}$ は $\boldsymbol{X}$ の分散共分散行列，λ_1 は $\boldsymbol{\Sigma}$ の最大固有値である。したがって，$\boldsymbol{\beta}$ は λ_1 に対応する固有ベクトルである。また，z_1 の分散は，

$$\mathrm{Var}(z_1)=\boldsymbol{E}(\boldsymbol{\beta}'\boldsymbol{X})^2=\boldsymbol{E}(\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta})=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}=\lambda_1\boldsymbol{\beta}'\boldsymbol{\beta}=\lambda_1 \tag{1.3}$$

となり，固有値に等しくなっている。

次に，z_1 と無相関な主成分 z_2 についても同様に，分散の最大化を考える。但し，ここでの制約条件は，$\mathrm{Cov}(z_1, z_2)=0$ と $\boldsymbol{\beta}'\boldsymbol{\beta}=1$ の2つになる。λ_2 を $\boldsymbol{\Sigma}$ の第2固有値とすれば，この場合も，$\boldsymbol{\Sigma}\boldsymbol{\beta}=\lambda_2\boldsymbol{\beta}$ という固有値問題を解けばよい。以下同様にして，$z_3\cdots\cdots z_p$ までの p 個の主成分が求められる。z_j は第 j 主成分と呼ばれる。

以上のように主成分分析は，p 個の互いに相関をもった変量を線形的に結合させ，互いに独立な総合的指標（＝主成分）にまとめる方法である。第1主成分は，分散の最大化という基準をもとに，p 個の変量の変動を最もよく表わした未知の合成変量である（p 個の変量と主成分との重相関係数の2乗和の最大化という基準からも，主成分は導出される)。また，X_j の分散の合計は，固有値の合計，すなわち p 個の主成分の分散の合計に等しいので，主成分分析は，観察データの分散という情報を考え，その情報の損失ができるだけ少なくなるように主成分を決定していく方法であるといえるだろう。つまり，X の変動（分散）を，できるだけ少数の主成分によって説明することが大きな目的である。$\lambda_k/\sum_{j=1}^{p}\lambda_j$ を第 k 主成分の寄与率といい，その主成分の説明力を表わす。もちろん，第1主成分の寄与率が最大であるが，できるだけ少数の主成分の累積寄与率が大きくなることが望ましい。

主成分分析の目的は，p 個の変量にみられる変動を要約的に記述し，次元を減少させることによって，複雑にからみあった変動の中から，その背後にある未知の共通の変動要因を抽出することにあるといえよう。

(2) CI の算式

以上のような性質をもつ主成分分析を CI のウェイトの算出に利用する前に，

CI の計算方法を示しておく。

CI は，景気動向指数のように個別指標の変化の方向のみに注目するのではなく，「変化率」に基づいて，景気を把握し，表現しようとする景気指標である。各個別指標は一定の基準によって選択され，それぞれ先行，一致，遅行の各系列に分類される。そして，その3系列について CI は次のように計算される。[6]

1. 対称変化率 c_{it} の計算

$$c_{it}=\frac{d_{it}-d_{i,t-1}}{d_{it}+d_{i,t-1}}\times 200 \qquad i=1,\cdots\cdots,p,\ t=2,\cdots\cdots N$$

d_{it}：季節調整済の各個別指標データ

p：個別指標数，N：観察期間数

但し，d_{it} が率の場合，$c_{it}=d_{it}-d_{i,t-1}$

2. 調整変化率 s_{it} の計算（c_{ti} の標準化）

$$s_{it}=\frac{c_{it}}{A_i} \qquad \text{但し，} A_i=\frac{\sum_{t=2}^{N}|c_{it}|}{N-1}$$

3. 平均変化率 r_t の計算（s_{it} の平均）

$$r_t=\frac{\sum_{i=1}^{p} w_i s_{it}}{\sum_{i=1}^{p} w_i}$$

w_i：第 i 指標に対するウェイト

4. 累積変化率 I_t の計算（r_t の累積）

$$I_t=I_{t-1}\times\frac{200+r_t}{200-r_t}$$

5. CI の計算（I_t の指数化）

$$CI_t=\frac{I_t}{\bar{I}}\times 100$$

$\bar{I}$：基準時の I_t

実際の CI の算式は，もっと複雑である。[7] しかし，基本的な部分はここに

示したもので十分であろう。平均変化率 r_t の計算ではウェイト w_i をつけているが，ウェイトをつけない経済企画庁の CI などの場合には，すべての $w_i=1$ としたものと考えればよい（すなわち，$r_t=\sum_{i=1}^{p} s_{it}/p$ として計算する)。しかし，各個別指標の変動の定義された景気に対する貢献度は異なっている，と考えた方が自然である。つまり，景気に大きな影響を与える個別指標と，そうでないものとがあると考えた方がよい。従って，その貢献度（影響力）が測定され，それに基づいたウェイトを付加して CI を計算すれば，ウェイトをつけないよりも，より正しく景気を表現することになるだろう。こうしたことから，本論文では，景気指標にウェイトをつけた方がよいという立場をとっている。(8)

(3)主成分分析による CI のウェイト

さて問題は，ウェイトとして，どのようにその貢献度を測定するかである。景気指標は，複雑な変動を示す個別指標を総合して景気を表現する。つまり，景気指標の作成において，景気とは，選択された個別指標の背後に潜む共通の変動要素の1つであるという操作的な定義がなされている。従って，その景気循環的変動要素が抽出され，それに対する各個別指標の影響力を測定できれば，ウェイトを算出することが可能である。

このように景気と景気指標，ならびにそのウェイトを考えれば，主成分分析の概念に一致する。すなわち，各個別指標に対して主成分分析を行い，それらの共通変動要素として主成分を抽出する。そして，抽出された主成分のうち景気と判断される主成分を発見し，その主成分（＝景気）に対する各個別指標の影響力を，対応する固有ベクトルの要素によって測定すればよい。つまり，固有ベクトル $\boldsymbol{\beta}$ を景気指標のウェイトとして用いることになる。

次に，一般的な景気指標ではなく，CI に対してはどのように主成分分析を行ったらよいかを考えよう。(9) もとのデータ d_{it} に対して主成分分析を行うことがまず考えられるが，CI のウェイトとしては，対称変化率 c_{it} や調整変化率 s_{it} に主成分分析を行った方が適切であると思われる。なぜならば，CI では

景気を変化率によって表現するので，各個別指標の共通変動要素というのは，c_{it} や s_{it} といった変化率の背後に存在すると考えられるからである。よって，c_{it} や s_{it} に主成分分析を行い，共通の変動要素を抽出すればよい。

注(5) 例えば，Anderson [2] pp. 453–455 などを見よ。

(6) Shiskin [29], Zarnowitz and Boschan [36] [37]，刈屋 [15] 127-129 ページ。

(7) この他にトレンドや振幅の調整などが行われる。

(8) ウエイトをつけないという立場は，単純平均の方がよいという考えからではなく，むしろ，妥当なウエイトが得られないという消極的な理由に基づいている。

ここで示す主成分分析以外のウエイトの算出方法としては，アメリカの CI におけるスコアリング・システムに基づく方法 (Zarnowitz and Boschan [36] [37] 参照) が有名である。その他の方法は，溝口・浜田 [21] 第4章，森 [22] 第4章などを参照のこと。

(9) 景気動向指数であれば，変化の方向を示す1，0データに主成分分析を適用すればよい。

2. 主成分分析における統計的推論

前節では，主成分分析によって CI のウェイトを算出することの有効性を示した。しかし，そうして得られたウェイトに対しては，記述統計の立場に基づいて経済学的な解釈が加えられるにすぎない。ウェイトの妥当性について，統計理論的な裏付けは与えられないのである。そこで本節では，主成分分析における統計的推論の諸結果をまとめ，ウェイトの検定への適用可能性を追求する。

(1)固有値・固有ベクトルの最尤推定量

以下では，母集団と標本を明確に区別するので，表1のように記号を定める。

表1 主成分分析における記号

	分散共分散行列	相関行列	固有値	固有ベクトル
母集団	$\boldsymbol{\Sigma}$	$\boldsymbol{P}$	$\lambda(\boldsymbol{\Lambda})$	$\boldsymbol{\beta}(\boldsymbol{B})$
標本	$\boldsymbol{S}$	$\boldsymbol{R}$	$l(\boldsymbol{L})$	$\boldsymbol{b}(\boldsymbol{B})$

$$\boldsymbol{\Lambda}=\lambda_i\boldsymbol{I},\quad \boldsymbol{L}=l_i\boldsymbol{I}$$

$$\boldsymbol{B}=[\boldsymbol{\beta}_1\cdots\boldsymbol{\beta}_p]\qquad \boldsymbol{B}=[\boldsymbol{b}_1\cdots\boldsymbol{b}_p]$$

$$\boldsymbol{S}=\frac{1}{N-1}\Sigma(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})'$$

ここで標本における主成分分析について，次の定理が成り立つ。[10]

定理 1　$X \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. $\lambda_1 > \lambda_2 > \cdots\cdots > \lambda_p$[11] とすれば

a．固有値 λ_i の最尤推定量は，$\boldsymbol{S}$ の p 個の異なった固有値を n/N 倍した $n/N \cdot l_i$ である（但し，$n=N-1$).

b．固有ベクトル $\boldsymbol{\beta}_i$ の最尤推定量は，標本固有ベクトル $\boldsymbol{b}_i$ である。

(2)固有値・固有ベクトルの分布

定理 2　$X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\lambda_1 > \lambda_2 > \cdots\cdots \lambda_p$, $n \to \infty$ とすれば，

a．l_i は $\boldsymbol{b}_i$ の各要素と独立に分布する。

b．$\sqrt{n}(l_i - \lambda_i)$ は，$N(0, 2\lambda_i^2)$ に従い，l_k　$(i \neq k)$ とは独立に分布する。

c．$\sqrt{n}(\boldsymbol{b}_i - \boldsymbol{\beta}_i)$ は，

$$N\left(0,\ \sum_{\substack{k=1 \\ k \neq i}}^{p} \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \boldsymbol{\beta}_k \boldsymbol{\beta}_k'\right)$$

という p 変量正規分布に従う。

d．$\boldsymbol{b}$ の要素 b_{kr} と b_{js} の共分散は，

$$-\frac{\lambda_k \lambda_j b_{kr} b_{js}}{n(\lambda_k - \lambda_j)^2} \qquad (k \neq j)$$

である。[12]

これらの定理は，後の固有値・固有ベクトルの検定に大きな役割を果たす。もちろん，これら定理からも，直接，いくつかの推論が可能である。例えば，定理 2-b によれば，$\sqrt{n}(l_i - \lambda_i)/\sqrt{2\lambda_i^2}$ は $N(0, 1)$ に従うので，信頼度 $1-\alpha$ の固有値の信頼区間は，$P(-z_{\frac{\alpha}{2}} \leq \sqrt{n/2} \cdot (l_i - \lambda_i)/\lambda_i \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$ より，

$$\frac{l_i}{1 + z_{\frac{\alpha}{2}}\sqrt{\dfrac{2}{n}}} \leq \lambda_i \leq \frac{l_i}{1 - z_{\frac{\alpha}{2}}\sqrt{\dfrac{2}{n}}} \tag{2.1}$$

と設定できる。これによって，固有値の精度について確率的に言及することが可能となる。

(3)固有ベクトルの検定

定理 2–c における $\sqrt{n}(\boldsymbol{b}_i-\boldsymbol{\beta}_i)$ の分散共分散行列を，$\boldsymbol{B}\boldsymbol{\varDelta}_i^2\boldsymbol{B}'=\boldsymbol{B}_i^*\boldsymbol{\varDelta}_i^{*2}\boldsymbol{B}_i^{*\prime}$ とする。ここで $\boldsymbol{\varDelta}_i$ は，対角要素が $\sqrt{\lambda_i\lambda_j}/(\lambda_i-\lambda_j)$ の p 次対角行列，$\boldsymbol{\varDelta}_i^*$ は $\boldsymbol{\varDelta}_i$ から i 行 i 列を除いた $(p-1)$ 次対角行列，$\boldsymbol{B}_i^*$ は $\boldsymbol{B}$ から i 行を除いた $(p-1)\times p$ 次の行列である。さらに，$\boldsymbol{h}_i=\boldsymbol{\varDelta}_i^*\boldsymbol{B}_i^{*\prime}\sqrt{n}(\boldsymbol{b}_i-\boldsymbol{\beta}_i)$ とおくと，定理 2 より $\boldsymbol{h}_i$ の極限分布は正規分布であり，

$$E(\boldsymbol{h}_i)=\boldsymbol{0} \tag{2.2}$$

$$\mathrm{Cov}(\boldsymbol{h}_i)=\boldsymbol{\varDelta}_i^{*-1}\boldsymbol{B}_1^{*-1}(\boldsymbol{B}_i^*\boldsymbol{\varDelta}_i^{*2}\boldsymbol{B}_i^{*\prime})\boldsymbol{B}_i^*\boldsymbol{\varDelta}_i^{*-1}=\boldsymbol{I}_{p-1}$$

となる。$\boldsymbol{h}_i\sim N(\boldsymbol{0},\boldsymbol{I})$ より，$\boldsymbol{h}'_i\boldsymbol{h}_i$ は，自由度 $(p-1)$ の χ^2 分布に従う。また，

$$\begin{aligned}\boldsymbol{h}'\boldsymbol{h}&=n(\boldsymbol{b}_i-\boldsymbol{\beta}_i)'\boldsymbol{B}_i^*\boldsymbol{\varDelta}_i^{*-2}\boldsymbol{B}_i^{*\prime}(\boldsymbol{b}_i-\boldsymbol{\beta}_i)\\&=n(\boldsymbol{b}_i-\boldsymbol{\beta}_i)'\left[\sum_{j=1}^{p}\boldsymbol{\beta}_j\left(\frac{\lambda_i}{\lambda_j}-2+\frac{\lambda_j}{\lambda_i}\right)\boldsymbol{\beta}'_j\right.\\&\qquad\left.-\boldsymbol{\beta}_i\left(\frac{\lambda_i}{\lambda_i}-2+\frac{\lambda_i}{\lambda_i}\right)\boldsymbol{\beta}_i'\right](\boldsymbol{b}_i-\boldsymbol{\beta}_i)\\&=n(\boldsymbol{b}_i-\boldsymbol{\beta}_i)'\left(\lambda_i\boldsymbol{\Sigma}^{-1}-2\boldsymbol{I}+\frac{1}{\lambda_i}\boldsymbol{\Sigma}\right)(\boldsymbol{b}_i-\boldsymbol{\beta}_i)\end{aligned} \tag{2.3}$$

である。ここで，$\boldsymbol{\Sigma}$, λ_i をそれぞれ $\boldsymbol{S}$, l_i でおきかえ，$\boldsymbol{h}'\boldsymbol{h}=\chi_0^2$ とすれば，

$$\begin{aligned}\chi_0^2&=n(\boldsymbol{b}_i-\boldsymbol{\beta}_i)'\left(l_i\boldsymbol{S}^{-1}-2\boldsymbol{I}+\frac{1}{l_i}\boldsymbol{S}\right)(\boldsymbol{b}_i-\boldsymbol{\beta}_i)\\&=n\left(l_i\boldsymbol{\beta}_i'\boldsymbol{S}^{-1}\boldsymbol{\beta}_i+\frac{1}{l_i}\boldsymbol{\beta}_i'\boldsymbol{S}\boldsymbol{\beta}_i-2\right)\end{aligned} \tag{2.4}$$

であり，χ_0^2 は漸近的に自由度 $(p-1)$ の χ^2 分布に従う。[13]

(2.4)を用いれば，$\boldsymbol{\beta}_i$ に関する推論が可能になる。信頼区間を設定することも考えられるが，ここでは $\boldsymbol{\beta}_i$ に関する仮説検定について示す。いま，帰無仮説・対立仮説をそれぞれ，

H_0 : $\boldsymbol{\beta}_i=\boldsymbol{\beta}_{i0}$

H_1 : $\boldsymbol{\beta}_i\neq\boldsymbol{\beta}_{i0}$ 　　但し，$\boldsymbol{\beta}_{i0}'\boldsymbol{\beta}_{i0}=1$

としよう。H_0 のもとで，(2.4) の $\boldsymbol{\beta}_i$ を $\boldsymbol{\beta}_{i0}$ におきかえて計算した χ_0^2 は，

自由度（$p-1$）の χ^2 分布に従う，よって，χ_0^2 を検定統計量に用いて，$\chi_0^2 \geq \chi_\alpha^2(p-1)$ であれば，H_0 は棄却される。

(4)固有値の検定

主成分分析では，固有値の相対的な大きさが，それに対応する主成分の有効性を示す指標になる。もちろん，それを直接利用した寄与率，または累積寄与率によって主成分の有効性を判断するのも，1つの便宜的方法であり，これまでの実証分析では，ほとんどそうした指標により，主成分の有効性を判断していた。しかし，定理1や定理2などで固有値に関する推論が可能になったので，それを利用した客観的な固有値の検定を行うことができる。

Anderson〔2〕では，固有値に関する3つの検定方法が示されている。ここでは，そのうち最も代表的な方法として，尤度比検定を応用した sphericity test を取り上げる。

いま，母集団における分散共分散行列の状態を，$\boldsymbol{\Sigma}=\boldsymbol{\Phi}+\sigma^2\boldsymbol{I}$ としよう。但し，$\boldsymbol{\Phi}$ は，階数が $m(<p)$ である正値半定符号行列である。この状態において，$\boldsymbol{\Sigma}$ の小さい方の（$p-m$）個の固有値は，すべて σ^2 に等しい。すなわち，$\boldsymbol{\Phi}$ が真の分散共分散行列を与え，σ^2 は誤差の分散である。従って，その（$p-m$）個の固有値は，$\lambda_{m+1}=\cdots\cdots=\lambda_p=\sigma^2$ となり，それらはまったく“random”であると考えられ，対応する主成分もデータの要約的記述には必要ないと判断できよう。

そこで，仮説を次のように設定する。

$H_0: \lambda_{m+1}=\cdots\cdots=\lambda_p$

$H_1: \lambda_{m+1},\cdots\cdots\lambda_p$ のすべてが等しいとは限らない。

H_0 が棄却されれば，（$p-m$）個の固有値には差があると判断される。H_0 が棄却されない場合，H_0 が正しいと解釈すれば，（$p-m$）個の固有値には差がない（たとえ，標本固有値 l_i に差がみられても，その差は random である）と判断され，重要な主成分は，はじめの m 個だけと結論づけられよう。

この仮説の検定統計量は，次のように求める。σ^2 の最尤推定量は，$l_{m+1},\cdots$

$\cdots l_p$ の算術平均 $\sum_{i=m+1}^{p} l_i/r$（但し，$r=p-m$）である。また，H_0 のもとでは l_i の算術平均と幾何平均は等しくなるので，

$$Q_r=\frac{\prod_{i=m+1}^{p} l_i}{\left(\frac{1}{r}\sum_{i=m+1}^{p} l_i\right)^r} \tag{2.5}$$

という尤度比を利用して検定を行う。但し，実際の検定統計量には，(2.5) のかわりに，$W_r=-n\log Q_r$ を用いる。W_r の計算式は

$$W_r=-n\sum_{i=m+1}^{p}\log l_i+nr\log\frac{\sum_{i=m+1}^{p} l_i}{r} \tag{2.6}$$

で与えられる。W_r の極限分布は，$X\sim N(\mu, \Sigma)$, $n\to\infty$ という仮定のもとで，自由度 $\phi=\frac{1}{2}(r-1)(r+2)$ の χ^2 分布であることがわかっている。[15] 従って，$W_r\geq\chi^2_\alpha(\phi)$ ならば H_0 は棄却され，$W_r<\chi^2_\alpha(\phi)$ ならば H_0 は棄却されず，先に示したような解釈がなされる。

この検定の大きな特徴は，(2.6)の W_r を計算する際に，標本観察値以外は必要としないことであり，与えられた標本において W_r の値は，r（またはm）の値のみに依存する。従って，分析に必要な主成分の数を客観的に決定できる。そこには，累積寄与率が，例えば80％越えるまでの主成分を取り上げる，といった曖昧さは存在しない。

(5)検定の手順

これまで示した以外の統計的推論を展開することも可能であるが，ウェイトの検定には以上のポイントで十分であろう。そこで，これらの結果を CI のウェイトの検定に具体的にどう適用したらよいかを示しておこう。

まず，固有値の検定を行う。ウェイトには固有ベクトルを用いるが，固有ベクトルの有効性は，あくまでもそれに対応する固有値の有効性に依存しているからである。[16] 固有値の検定において，与えられた標本で H_0 が棄却されるかどうかは，m（または r）の値に依存する。そこで，mの値を 1，2, …… pと変化させ，それぞれの W_r について有意性を確かめる。はじめのうちは H_0 が

棄却されることが予想されるが，あるmの値から H_0 が棄却されなくなったら，その直前までの主成分を必要な主成分とみなせばよい。[17] H_0 が棄却されないということは，対応するr個の主成分が重要でないことを示すからである。例えば，$m=2$ までは H_0 が棄却され，$m=3$ 以下で H_0 が棄却されなければ，第3主成分以下は無視して第1，第2主成分がデータの変動を説明する要因であると判断し，利用すればよい。

こうした検定方法は，できるだけ少ない次元によるデータの記述という主成分分析の目的に合っている。もちろん，ある主成分だけを対象にmを固定してもかまわない。景気指標においては，景気と判断される主成分に対応する固有値について W_r で検定してもよいし，$m=1\cdots\cdots p$ と変化させて，有意と判断される主成分のなかに景気を示す主成分が含まれるかをみてもよい。

次に，景気を示す主成分が有意と判断されれば，その固有ベクトルを検定する。ここで問題となるのは，(2.4)に代入する $\boldsymbol{\beta}_i=\boldsymbol{\beta}_{i0}$ の決め方である。CI のウェイトとして適切かどうかを検定するためには，$\boldsymbol{\beta}_{i0}$ のすべての要素が等しいという仮説をたてればよい。$\boldsymbol{\beta}_{i0}'\boldsymbol{\beta}_{i0}=1$ という制約条件より $\boldsymbol{\beta}_{i0}$ の各要素を$1/\sqrt{p}$とし，

$$H_0:\boldsymbol{\beta}_i=\boldsymbol{\beta}_{i0}=\left[\frac{1}{\sqrt{p}}\cdots\cdots\frac{1}{\sqrt{p}}\right]$$

$H_1:\boldsymbol{\beta}_i$ のすべての要素が等しいとは限らない。

という仮説をたてることになる。H_0 が正しければ，等ウェイトと判断され，単純平均がよいことになる。H_0 が棄却されれば，ウェイトとして用いる固有ベクトルの各要素は異なっていると判断される。すなわち，異なったウェイトをつけること（加重平均）に対して，統計理論的な裏付けが与えられることになる。もちろん，ウェイトには $\boldsymbol{b}_i$ を用いればよい。

注(10) Anderson [2] pp. 460–461.

(11) λ に重複 (multiplicity) があっても，若干の修正により定理1は成立する。その場合も含めた定理1の証明は，Anderson [1] を参照のこと。

(12) 奥野他 [23] 191–192 ページ。証明は，Girshick [10] [11]，Bartlett [3] [4]，Lawley [18] などに示されている。

(13) Anderson [2] pp. 469–471.

(14) Anderson [2] pp. 473–477. ここで取り上げる以外の2つの検定も，検定の目的は同じである。しかし，H_0 の設定に一定の水準を与えねばならず，やや客観性に欠ける。

(15) 証明は，Anderson [1] pp. 130–133 を参照のこと。

(16) Anderson [2] p. 471.

(17) もちろん，すべての$\boldsymbol{m}$に対して，H_0 が棄却されたり，棄却されなかったりする可能性もある。しかし，一般的にいって$\boldsymbol{m}$が小さければ H_0 は棄却されやすい。

3. 固有値検定の問題点

これまで示してきたような主成分分析における推論は，実証的にはほとんど利用されてない。その理由は，第1に，主成分分析には得られた主成分の解釈とデータとの整合性の検討，及びその応用といった記述統計的側面が強いからである。第2は，先に示したような検定の具体的方法が示されなかった点も含めて，推論のもつ方法論上の問題点が多いことである。こうした問題点について，最も重要な固有値の検定を中心に，以下で考察してみよう。

(1)W_r の性質

固有値の検定統計量 W_r は，一般に$\boldsymbol{n}$の大きさに左右されやすい。相加平均は相乗平均より大きいので，(2.6) の第2項は第1項より$\boldsymbol{n}$に敏感である。$\boldsymbol{n} \rightarrow \infty$と仮定しているから，この検定では，$H_0$ が棄却されやすいといえよう。

(2)仮説のたて方

統計的仮説検定においては，棄却したい仮説を H_0 にするべきで，H_0 が棄却されないからといって，H_0 を正しいと判断することは危険である。しかし，固有値の検定では，H_0 が棄却されない場合にのみ，第$\boldsymbol{m}$主成分までを有効だと判断せざるを得ない。これは，H_0 のたて方に問題があるからである。だからといって，H_0 と H_1 を入れかえると，検定統計量が得られなくなり，検定は不可能になる。したがって，W_r 以外の情報も利用して，慎重な判断を下す必要がある。

(3)正規性の仮定について

固有値の検定を展開するのには，3つの仮定が必要である。そのうち，$\lambda_1 > \cdots\cdots > \lambda_p$ と $n \to \infty$ という仮定は，主成分分析においてそれほど非現実的ではない。しかし，もう1つの $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ という仮定は，現実的には成立しにくい。特に，経済時系列データなどでは，標本の独立性・正規性の仮定は疑わしいものになる。[18] そこで，統計的推論の基礎となるこの正規性の仮定を検討する必要がある。それには，次の2つのアプローチが考えられる。

第1は，正規性の仮定をはずして，非正規母集団に基づいた推論を行う方法である。[19] W_r に関して示すと，母集団に4次キュムラント κ_4 の存在を仮定すれば，正規母集団でなくても，$W_r^* = W_r/(\hat{\kappa}+1)$ が自由度 $\frac{1}{2}(r+2)(r-1)$ の χ^2 分布をすることがわかっている[20]（但し，$\kappa_4 = 3\kappa$, $\hat{\kappa}$ は κ の一致推定量)。よって，W_r^* を用いて先と同様の固有値の検定ができる。

また別の方法として，l_i や $\boldsymbol{b}_i$，さらに W_r などが，正規性の仮定がなくても安定的になるようなロバストな方法も考えられている。[21] 実際には，l_i や $\boldsymbol{b}_i$ の計算の基礎となる $\boldsymbol{S}$ や $\boldsymbol{R}$ を，M 推定量，trimmed-mean などでロバスト化する。

第2のアプローチは，$\boldsymbol{X}$ が $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ からの標本かどうかを検定する方法である。正規性の検定には，正規確率紙への確率プロットや検定統計量による方法などがある。代表的なものとして，$\sqrt{b_1}$（歪度），b_2（尖度），χ^2 検定，Cramér-von Mises 検定，Kolmogorov-Smirnov 検定，Shapiro-Wilk の W 統計量を用いた検定などがあげられる。[22] しかし，これらはいずれも1変量正規分布の検定である。多変量正規分布の検定には，上のような検定統計量を一般化して用いればよい[23]が，かなり複雑である。p 個の変量それぞれについて確率プロットをする方法や検定統計量を計算する方法もあるが，それは周辺分布の正規性しか確かめることができない。変数間の相関を考慮していないことが問題で，結合分布の正規性の検定にはならないのである。

そこでこの問題点を改良した方法に，主成分分析を利用した方法がある。[24] $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ であれば，主成分 $\boldsymbol{\beta}_i'\boldsymbol{X}$ は互いに独立に正規分布をするので，

得られたp個の主成分を，確率プロットなどで正規性の検定を行えば，X の正規性が確められる。これは主成分が独立であるという性質を利用した興味深い検定方法である。

これらのどの検定がよいかは，まだはっきりと結論がでていない。いずれにせよ，$X \sim N(\mu, \Sigma)$ がいえれば，W_r をそのまま用いて検定を行い，X が正規分布に従わないと判断されれば，X に適当な変数変換(25)を行い，正規性を確保する必要がある。実証分析においては，統計的推論の前提となる正規性の仮定に対して，何らかの検討を加えなければならないだろう。

注(18) Bolch and Huang [5] p. 250.
(19) Davis [7], Fujikoshi [9], Waternaux [34] [35] など。
(20) Waternaux [35] pp. 327–330.
(21) Devlin et al. [8]
(22) 岩田 [14] 第8章, Shapiro et al. [26], Shapiro and Wilk [27], 柴田 [28] 第7章, Srivastava and Carter [31] chap. 3 などを参照のこと。特に Shapiro et al. [26] では，各統計量を比較している。
(23) Cox and Small [6], Malkovich and Afifi [20], Small [30]. $\sqrt{b_1}$ や b_2 に関しては，Mochado [19] などにくわしい。
(24) Srivastava and Carter [31] pp. 286–288.
(25) Srivastava and Carter [31] pp. 68–71.

4. 数値例

最後に数値例を示そう。経済企画庁が発表している CI の一致系列のうち9系列(表2)を用いて計算する。対象期間は，昭和50年1月から昭和53年12月までの48か月間である(26)（データは『経済変動観測資料年報』昭和60年度による)。

表3は，このデータに主成分分析を行った結果である。(27) W_r の変化をみると，第2主成分から第3主成分へかわるとき有意でなくなる。(28) 従って，第3主成分以下の主成分は無視し得ると解釈できる。しかし，第2主成分までの累積寄与率も43%と低く，2つの主成分でデータを要約できるかは疑問である。そこで，有意と判断された2つの主成分の固有ベクトルを，表4によって確かめてみよう。表4によれば，第2主成分が各変数に共通な変動要素（＝景気)

表2　CI 採用系列一覧（一致系列）

一致系列

1. 生産指数（鉱工業）　（季）
2. 生産者出荷指数（鉱工業）　（季）
3. 稼働率指数（製造業）　（季）
4. 原材料消費指数（製造業）　（季）
5. 大口電力使用量　（季）
6. 輸入数量指数　（季）
7. 建築着工床面積（鉱工業）　（季）
8. 月間有効求人数（学卒を除く・全産業）*　前年同月比
9. 百貨店販売額*　前年同月比

（季）は，季節調整済系列を示す
*　は，率の系列を示す

表3　主成分分析の結果

主成分	固有値	寄与率	累積寄与率	W_r	d. f
1	2.186	.2429	.2429	79.59*	44
2	1.726	.1978	.4347	56.00*	35
3	1.218	.1353	.5700	37.03	27
4	1.008	.1120	.6820	27.60	20
5	.8836	.0982	.7802	20.52	14
6	.7451	.0828	.8630	13.46	9
7	.5076	.0564	.9194	6.641	5
8	.4758	.0529	.9723	4.811	2
9	.2496	.0277	1.000	0	0

* は5％有意を示す

表4　固有ベクトルの要素

変数	第 1 主成分	第 2 主成分	第 3 主成分
1	−.3696	.4029	−.2433
2	.0580	.5126	−.0799
3	−.3822	.3309	.3282
4	−.2160	.2632	.5701
5	.3553	.3885	.3182
6	−.3185	−.2068	.1612
7	−.2861	.3184	−.5757
8	.5217	.2645	.1651
9	.2926	.1767	.0833

を表わしていると解釈できる。なぜならば，輸入数量指数（変数6）を除いて，すべての個別指標に対応する固有ベクトルの要素がプラスであり，第2主成分が，各指標の変化の方向に等しい共通の変動を示すと考えられるからである。もちろん，ここで用いられている各個別指標が，共通して景気をよく表わすものであるという前提で議論を進めている。

次に景気を示すと思われた第2主成分の固有ベクトルの検定を行う。$p=9$ より，仮説は

$$H_0:\boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}=\left[\frac{1}{3}\cdots\cdots\frac{1}{3}\right]$$

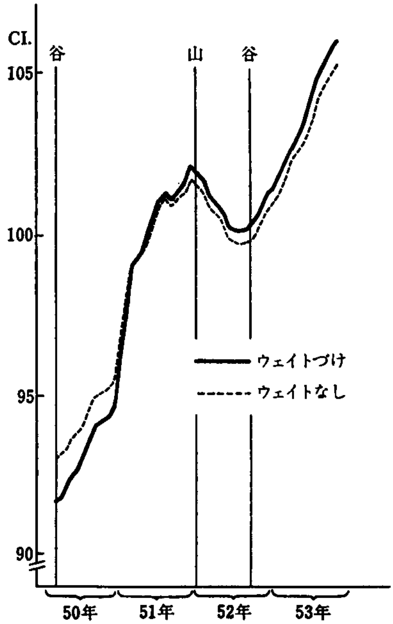
$$H_1:\boldsymbol{\beta}_2\neq\boldsymbol{\beta}_{20}$$

である。(2.4) に基づいて χ_0^2 を計算すると 43.31 になる。$\chi_{0.05}^2(8)=15.51$なので H_0 は棄却され，$\boldsymbol{\beta}_2$ の各要素は等しくないと判断できる。このことからウェイトに $\boldsymbol{b}_2$ を用いて，CI を加重平均によって計算することの妥当性が確かめられたといえよう。

図1は，$\boldsymbol{b}_2$ によってウェイトをつけたCIとウェイトをつけないCIの動きを示している。ともに景気の転換点（景気の山，谷）との対応は良好で大きな違いはなく，ウェイトづけを否定する材料はない。むしろ，ウェイトをつけた CI の方が景気の循環的変動を強調しており，CI による景気循環を把握しやすくしているといえよう。

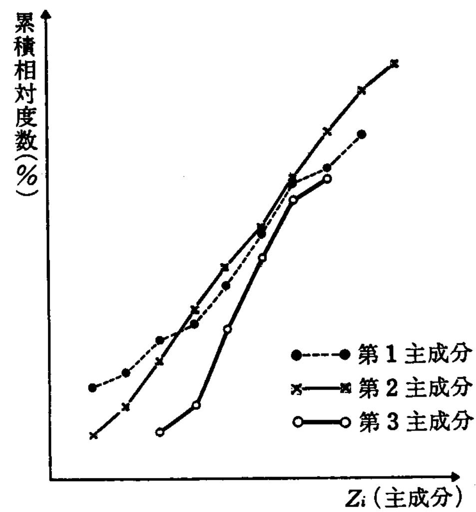
最後に母集団の正規性を検討しよう。ここでは，先に述べた主成分を確率プロットする方法を用いてみる。図2は，第1～第3主成分を正規確率紙にプロットしたものである。これだけでは完全に結論を出せないが，$\boldsymbol{X}$はそれほど正規分布から乖離していないように思われる。主成分分析

図1　CI の動き（3か月移動平均値）



『経済変動観測資料年報』昭和60年度のデータより算出

図2　主成分の確率プロット（正規確率紙）



を原データではなく，変化率に対して行ったことが，正規分布に近いと判断させる原因の1つであろう。

これらの結果から，固有値，固有ベクトルの検定を受け入れ，$\boldsymbol{b}_2$ を CI のウェイトとして用いることは適切であるといえる。

注(26) この期間を選んだ理由は，比較的短い期間に3つの転換点が含まれ，CI のパフォーマンスをみるのに好都合だからである。あまり長い期間をとると，構造変化などが生じるので，分析上，好ましくない。

(27) 主成分分析にかけたのは，原データでなく調整変化率 s_{it} である。

(28) 連続的な検定では，各検定の有意水準と全体としての有意水準とは一致しない。ここでは，連続的な検定であるが，各々の検定は独立であると仮定して，一定の有意水準を与えている。

おわりに

以上，主成分分析の CI のウェイトへの適用可能性を考察してきた。とりわけ，ウェイトに関して統計的仮説検定が可能であるということは，かなり有意義であろう。従来の経済指数のウェイトは，物価指数に代表されるように，[29] 経験的または経済学的な基準に基づいて算出されている。これに対して，主成分分析によって算出されたウェイトには，本論文で示したように統計理論上の裏付けを与えることができる（ここでは行っていないが，ウェイトに対する経済学的な解釈も可能である）。

しかし，検定の存在が主成分分析を用いる直接の理由ではない。基本的には，景気指標における景気の概念と主成分分析とがうまく対応しているからであり，あくまでも検定はそれを統計理論的に検証するための方法にすぎないのである。但し，このように経済分析へ統計理論を適用する場合には，正規性の仮定の検討など，十分な注意が払われねばならない。

注(29) 消費者物価指数では，消費支出に対する各品目の支出の万分比が，その品目のウエイトになる。

参考文献

[1] Anderson. T. W., "Asymptotic Theory for Principal Component Analysis", *Annals of Mathematical Statistics,* vol. 34, 1963, pp, 122–148.

[2] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis,* 2 nd ed., John Wiley & Sons, 1984, 675 pp.

[3] Bartlett, M. S., "The Effect of Standardization of a χ^2-approximation in Factor Analysis", *Biometrika,* vol. 38, 1951, pp. 337–344.

[4] Bartlett, M. S., "A Note on the Multiplying Factors for Various χ^2-approximations", *Journal of the Royal Statistical Society (Series B),* vol. 16, 1954 pp. 296–298.

[5] Bolch, B. W. and Huang, C. J., *Multivariate Statistical Methods for Business and Economics,* Prentice-Hall, Inc., 1974, 329 pp. (中村慶一訳『応用多変量解析』森北出版，1976年，286ページ)

[6] Cox, D. R. and Small, J. H., "Testing Multivariate Normality", *Biometrika,* vol. 65, 1978, pp. 263–272.

[7] Davis, A. W., "Asymptotic Theory for Principal Component Analysis: Non-Normal Case", *Australian Journal of Statistics,* vol. 19, 1977, pp. 206–212.

[8] Devlin, S. J., Gnanadesikan, R. and Kettenring J. R., "Robust Estimation of Dispersion Matrices and Principal Components", *Journal of the American Statistical Association,* vol. 76, 1981, pp. 354–362.

[9] Fujikoshi, Y., "Asymptotic Expansion for the Distribution of the Sample Roots under Nonnormality", *Biometrika,* vol. 67, 1980, pp. 45–51.

[10] Girshick, M. A., "Principal Components", *Journal of the American Statistical Association,* vol. 31, 1936, pp. 519–528.

[11] Girshick, M. A., "On the Sampling Theory of Roots of Determinantal Equations", *Annals of Mathematical Statistics,* vol. 10, 1939, pp. 203–224.

[12] Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology,* vol. 24, 1933, pp. 417–441 and 498–520.

[13] Hotelling, H., "Relations between Two Sets of Variates", *Biometrika,* vol. 28, 1936, pp. 321–377.

[14] 岩田暁一『計量経済学』，有斐閣，1982年，364ページ．

[15] 刈屋武昭『計量経済分析の考え方と実際』，東洋経済新報社，1986年，189ページ．

[16] Kshirsagar, A. M., "The Goodness of Fit of a Single (non-isotropic) Hypothetical Principal Component", *Biometrika,* vol. 48, 1961, pp. 397–407.

[17] Kshirsager, A. M., *Multivariate Analysis,* Marcell Decker, 1972, 534 pp.

[18] Lawley, D. N., "Tests of Significance for the Latent Roots of Covariance

and Correlation Matrices", *Biometrika*, vol. 43, 1956, pp. 128-136.

[19] Machado, S. G., "Two Statistics for Testing for Multivariate Normality", *Biometrika*, vol. 70, 1983, pp. 713-718.

[20] Malkovich, J. F. and Afifi, A. A., "On Tests for Multivariate Normality", *Journal of the American Statistical Association*, vol. 68, 1963, pp. 176-179.

[21] 溝口敏行・浜田宗雄『経済時系列の分析』, 勁草書房, 1969年, 432ページ.

[22] 森一夫『日本の経済予測』, 東洋経済新報社, 1976年, 227ページ.

[23] 奥野忠一他『多変量解析法』〈改訂版〉, 日科技連, 1981年, 430ページ.

[24] Pearson, K., "On Lines and Planes Closest Fit to Systems of Points in Space", *Philosophical Magazine*, vol. 2, 1901, pp. 559-572.

[25] Rao, C. R., "The Use and Interpretation of Principal Component Analysis in Applied Research", *Sankhya A*, vol. 26, 1964, pp. 329-358.

[26] Shapiro, S. S., Wilk, M. B. and Chen, H. J., "A Comparative Study of Various Tests for Normality", *Journal of the American Statistical Association*, vol. 63, 1968, pp. 1343-1372.

[27] Shapiro, S. S. and Wilk, M. B., "An Analysis of Variance Test for Normality (complete samples)", *Biometrika*, vol. 52, 1965, pp. 591-611.

[28] 柴田義貞『正規分布—特性と応用』, 東大出版会, 1981年, 307ページ.

[29] Shiskin, J., *Signals of Reccession and Recovery*, National Bureau of Economic Research Occasional Paper 77, 1961, 191 pp.

[30] Small, J. H., "Plotting Squared Radii", *Biometrika*, vol. 65, 1978, pp. 657-658.

[31] Srivastava, M. S. and Carter, E. M., *An Introduction to Applied Multivariate Statistics*, North-Holland, 1983, 394 pp.

[32] 竹内啓『数理統計学』, 東洋経済新報社, 1963年, 373ページ.

[33] 田原昭四『景気変動と日本経済』, 東洋経済新報社, 1983年, 270ページ.

[34] Waternaux, C. M., "Asymptotic Distribution of the Sample Roots for a Non-Normal Population", *Biometrika*, vol. 63, 1976, pp. 639-645.

[35] Waternaux, C. M., "Principal Components in the Nonnormal Case: The Test of Equality of Q Roots", *Journal of Multivariate Analysis*, vol. 14, 1984, pp. 323-335.

[36] Zarnowitz, V. and Boschan C., "Cyclical Indicators: An Evaluation and New Leading Indexes", *Business Conditions Digest*, March 1975, pp. v-xxii.

[37] Zarnowitz, V. and Boschan C., "New Composite Indexes of Coincident and Lagging Indicators", *Business Conditions Digest*, Nov. 1975, pp. v-xxiv.

1986. 9. 30 脱稿

(博士後期課程第1年度生・統計学　佐竹元一郎教授研究指導)